

Penerapan Metode *Exploratory Data Analysis* Untuk Mengidentifikasi Faktor Peningkatan Pendapatan Pada Supermarket XYZ

Badriah Nursakinah¹, Hadi Zakaria^{2*}, Cintya Laura Melagro³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

Email : dosen02779@unpam.ac.id, dosen00274@unpam.ac.id

(* : coresponding author)

Abstrak - Perkembangan teknologi digital mendorong perusahaan ritel untuk memanfaatkan data penjualan sebagai dasar pengambilan keputusan bisnis. Penelitian ini bertujuan menerapkan *data mining* melalui metode *Exploratory Data Analysis* (EDA) untuk mengidentifikasi produk yang paling diminati, kategori produk dengan keuntungan tertinggi, wilayah dengan kontribusi penjualan terbesar, serta karakteristik konsumen berdasarkan gender pada Supermarket XYZ. Data dikumpulkan melalui observasi, wawancara, studi kepustakaan, dan dokumentasi data transaksi penjualan. Tahapan analisis mencakup pemeriksaan dan pembersihan data, pengelompokan data, analisis statistik deskriptif, serta visualisasi untuk menemukan pola dan kecenderungan penjualan. Hasil penelitian menunjukkan bahwa kategori *Electronic Accessories* merupakan produk yang paling banyak diminati dengan persentase pembelian sebesar 17,62%. Kategori *Food and Beverages* menghasilkan keuntungan tertinggi, yaitu sekitar Rp50.000, sedangkan kategori *Health and Beauty* memperoleh keuntungan terendah sekitar Rp10.000. Konsumen perempuan tercatat melakukan pembelian lebih banyak dibandingkan konsumen laki-laki, meskipun perbedaannya tidak signifikan. Penerapan EDA mampu menghasilkan informasi mengenai pola penjualan dan karakteristik konsumen yang dapat digunakan sebagai dasar penyusunan strategi persediaan, pemasaran, promosi, dan segmentasi pelanggan untuk mendukung peningkatan pendapatan Supermarket XYZ.

Kata kunci: *Data Mining, Exploratory Data Analysis, Pendapatan, Penjualan, Supermarket.*

Abstract— The development of digital technology has encouraged retail companies to utilize sales data as a basis for business decision-making. This study aims to apply data mining through the *Exploratory Data Analysis* (EDA) method to identify the most popular products, product categories with the highest profits, regions with the largest sales contributions, and consumer characteristics based on gender at XYZ Supermarket. Data were collected through observation, interviews, literature review, and sales transaction documentation. The analytical stages included data examination and cleaning, data grouping, descriptive statistical analysis, and visualization to identify sales patterns and trends. The results indicate that *Electronic Accessories* were the most popular product category, accounting for 17.62% of purchases. The *Food and Beverages* category generated the highest profit of approximately IDR 50,000, while the *Health and Beauty* category recorded the lowest profit of approximately IDR 10,000. Female consumers made more purchases than male consumers, although the difference was not significant. The application of EDA successfully generated information on sales patterns and consumer characteristics, which can serve as a basis for developing inventory, marketing, promotional, and customer segmentation strategies to support revenue growth at XYZ Supermarket.

Keywords: *Data Mining, Exploratory Data Analysis, Revenue, Sales, Supermarket.*

1. PENDAHULUAN

Perkembangan teknologi digital telah menjadikan data sebagai salah satu aset penting bagi perusahaan dalam mendukung kegiatan operasional dan pengambilan keputusan. Pertumbuhan jumlah data yang semakin besar mendorong perusahaan untuk mengelolanya secara optimal agar dapat menghasilkan informasi yang bermanfaat bagi pengembangan strategi bisnis. Salah satu pendekatan yang dapat digunakan adalah *data mining*, yaitu proses untuk menemukan informasi, hubungan, kecenderungan, dan pola tertentu dari sekumpulan data yang berukuran besar.

Pemanfaatan *data mining* menjadi semakin penting dalam industri ritel karena kegiatan transaksi menghasilkan data penjualan dalam jumlah besar setiap harinya. Data tersebut memuat berbagai informasi, seperti kategori produk, jumlah pembelian, keuntungan, lokasi cabang, dan karakteristik pelanggan. Namun, data transaksi belum dapat memberikan manfaat secara maksimal apabila hanya disimpan tanpa melalui proses pengolahan dan analisis. Oleh karena itu, diperlukan

metode analisis yang mampu mengubah data mentah menjadi informasi yang relevan sebagai dasar penyusunan strategi bisnis.

Supermarket XYZ merupakan perusahaan ritel yang beroperasi di tengah persaingan bisnis yang ketat. Kondisi tersebut menuntut perusahaan untuk memahami pola penjualan dan perilaku konsumennya agar dapat mempertahankan daya saing serta meningkatkan pendapatan. Beberapa persoalan yang perlu diperhatikan meliputi produk yang paling banyak diminati, kategori produk yang memberikan keuntungan terbesar, cabang yang memberikan kontribusi pendapatan tertinggi, serta perbedaan perilaku pembelian berdasarkan gender. Informasi tersebut diperlukan untuk mendukung pengelolaan persediaan, penentuan prioritas produk, penyusunan program promosi, dan segmentasi pelanggan.

Penelitian ini menerapkan metode *Exploratory Data Analysis* (EDA) untuk mengeksplorasi data transaksi penjualan Supermarket XYZ. EDA merupakan pendekatan analisis yang digunakan untuk memahami karakteristik dan struktur data, mendeteksi data yang tidak wajar, serta mengidentifikasi pola dan hubungan antarvariabel. Proses tersebut dilakukan melalui analisis statistik deskriptif dan penyajian visual dalam bentuk tabel, grafik, maupun diagram. Visualisasi data membantu menyederhanakan informasi yang kompleks sehingga hasil analisis lebih mudah dipahami oleh pihak manajemen.

Melalui penerapan EDA, data penjualan dapat dianalisis berdasarkan kategori produk, tingkat keuntungan, lokasi cabang, dan gender pelanggan. Analisis produk yang paling diminati dapat membantu perusahaan menentukan prioritas persediaan dan strategi promosi. Identifikasi kategori dengan keuntungan tertinggi dapat menjadi dasar dalam mengoptimalkan portofolio produk. Sementara itu, analisis kontribusi setiap cabang memungkinkan manajemen mengevaluasi perbedaan kinerja penjualan dan merumuskan strategi untuk meningkatkan kinerja cabang yang belum optimal.

Selain menganalisis produk dan cabang, EDA juga dapat digunakan untuk memahami karakteristik pelanggan. Analisis berdasarkan gender memberikan gambaran mengenai kelompok konsumen yang lebih dominan dalam melakukan pembelian. Informasi tersebut dapat dimanfaatkan untuk menyusun strategi pemasaran yang lebih terarah dan sesuai dengan karakteristik pelanggan. Dengan demikian, perusahaan dapat meningkatkan relevansi promosi, kepuasan pelanggan, dan peluang terjadinya pembelian berulang.

Keakuratan hasil EDA sangat dipengaruhi oleh kualitas data yang digunakan. Data yang tidak lengkap, tidak konsisten, atau mengandung kesalahan dapat menghasilkan interpretasi yang kurang tepat. Oleh karena itu, penelitian ini juga mencakup pemeriksaan, pembersihan, dan pengelompokan data sebelum proses analisis dilakukan. Tahapan tersebut diperlukan agar pola yang dihasilkan dapat menggambarkan kondisi penjualan secara lebih akurat dan mendukung pengambilan keputusan yang tepat.

Berdasarkan uraian tersebut, penelitian ini bertujuan menerapkan *data mining* menggunakan metode *Exploratory Data Analysis* untuk mengidentifikasi faktor-faktor yang berpotensi mendukung peningkatan pendapatan Supermarket XYZ. Secara khusus, penelitian dilakukan untuk mengidentifikasi produk yang paling diminati, menentukan kategori produk dengan keuntungan tertinggi, mengetahui cabang dengan kontribusi pendapatan terbesar, serta menganalisis perilaku pembelian konsumen berdasarkan gender. Hasil penelitian diharapkan dapat menjadi dasar bagi manajemen dalam menyusun strategi persediaan, pemasaran, promosi, dan segmentasi pelanggan secara lebih efektif.

2. METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan menerapkan teknik *data mining* melalui metode *Exploratory Data Analysis* (EDA). Pendekatan tersebut digunakan untuk mengeksplorasi data transaksi penjualan Supermarket XYZ dan mengidentifikasi pola-pola yang berkaitan dengan peningkatan pendapatan. Analisis difokuskan pada tingkat minat konsumen

terhadap kategori produk, keuntungan setiap kategori, kontribusi penjualan berdasarkan cabang atau kota, serta karakteristik pembelian berdasarkan gender pelanggan.

2.1 Pengumpulan Data

Data utama dalam penelitian ini berupa data transaksi penjualan Supermarket XYZ. Data tersebut mencakup informasi mengenai nomor transaksi, cabang, kota, kategori produk, jenis pelanggan, gender, harga satuan, jumlah produk, total transaksi, tanggal pembelian, metode pembayaran, dan keuntungan kotor. Untuk melengkapi serta mendukung proses analisis, pengumpulan data dilakukan melalui beberapa teknik berikut.

a. Observasi

Observasi dilakukan untuk memperoleh gambaran mengenai kegiatan penjualan, pola pembelian pelanggan, penempatan produk, pelaksanaan promosi, dan kondisi operasional Supermarket XYZ. Hasil observasi digunakan sebagai informasi pendukung dalam menginterpretasikan pola yang ditemukan pada data transaksi.

b. Wawancara

Wawancara dilakukan kepada pihak manajemen atau karyawan yang memahami kegiatan penjualan dan pengelolaan data supermarket. Wawancara bertujuan memperoleh informasi mengenai data yang tersedia, strategi pemasaran, pengelolaan persediaan, karakteristik pelanggan, dan kendala yang dihadapi dalam meningkatkan pendapatan.

c. Dokumentasi

Dokumentasi dilakukan dengan menghimpun data transaksi penjualan yang digunakan sebagai sumber utama analisis. Data tersebut dikelompokkan berdasarkan kategori produk, keuntungan, lokasi cabang, dan gender konsumen sesuai dengan tujuan penelitian.

d. Studi kepustakaan

Studi kepustakaan dilakukan dengan mempelajari buku, artikel ilmiah, jurnal, dan sumber lain yang relevan dengan *data mining*, EDA, visualisasi data, perilaku konsumen, dan strategi peningkatan pendapatan pada industri ritel.

2.2 Tahapan Pengolahan dan Analisis Data

Tahapan penelitian disusun secara sistematis agar data yang digunakan memiliki kualitas yang memadai dan hasil analisis dapat diinterpretasikan secara tepat.

a. Penetapan tujuan analisis

Tahap awal dilakukan dengan menentukan tujuan dan pertanyaan analisis. Penelitian diarahkan untuk menjawab empat pertanyaan utama, yaitu kategori produk yang paling diminati, kategori dengan keuntungan tertinggi, cabang atau kota dengan kontribusi pendapatan terbesar, dan gender yang paling dominan dalam melakukan pembelian.

b. Persiapan perangkat analisis

Pengolahan data dilakukan menggunakan bahasa pemrograman Python dengan pustaka yang sesuai, seperti Pandas untuk pengelolaan data serta Matplotlib dan Seaborn untuk visualisasi. Tableau dapat digunakan untuk menyajikan hasil analisis dalam bentuk grafik dan *dashboard* interaktif agar informasi lebih mudah dipahami.

c. Pemuatan data

Dataset dimuat ke dalam lingkungan analisis sesuai dengan format sumbernya, seperti CSV atau Excel. Pada tahap ini dilakukan pemeriksaan awal terhadap jumlah baris dan kolom, nama atribut, tipe data, serta kesesuaian struktur dataset dengan kebutuhan penelitian.

d. Pembersihan data

Pembersihan data dilakukan untuk meningkatkan kualitas dan konsistensi data. Tahapan ini meliputi:

- Pemeriksaan dan penanganan data kosong atau *missing value*;
- Penghapusan data duplikat;
- Penyesuaian tipe data, khususnya pada tanggal dan variabel numerik;
- Penyeragaman penulisan kategori produk, cabang, kota, dan gender;
- Pemeriksaan nilai yang tidak wajar atau *outlier*; dan
- Penghapusan atribut yang tidak relevan dengan tujuan analisis.

Penanganan data kosong disesuaikan dengan karakteristik dan proporsi data yang hilang. Data dapat dihapus apabila jumlahnya sedikit dan tidak memengaruhi keseluruhan analisis atau diisi menggunakan nilai yang sesuai, seperti rata-rata, median, atau modus.

e. Eksplorasi data

EDA dilakukan melalui analisis univariat dan bivariat. Analisis univariat digunakan untuk menggambarkan distribusi setiap variabel, sedangkan analisis bivariat digunakan untuk membandingkan dua variabel yang berkaitan. Proses eksplorasi meliputi:

- Menghitung frekuensi dan persentase transaksi setiap kategori produk;
- Menjumlahkan keuntungan berdasarkan kategori produk;
- Membandingkan total transaksi berdasarkan cabang atau kota;
- Menghitung jumlah pembelian berdasarkan gender konsumen; dan
- Mengidentifikasi pola, kecenderungan, serta anomali pada data penjualan.

Persentase pembelian setiap kategori produk dapat dihitung menggunakan rumus berikut:

$$P_i = \frac{f_i}{N} \times 100\%$$

Keterangan:

- P_i = persentase pembelian kategori ke- i ;
- f_i = jumlah transaksi kategori ke- i ; dan
- N = jumlah seluruh transaksi.

Sementara itu, keuntungan setiap kategori ditentukan dengan menjumlahkan nilai keuntungan seluruh transaksi pada kategori yang sama:

$$K_i = \sum_{j=1}^{n_i} k_{ij}$$

Keterangan:

- K_i = total keuntungan kategori ke- i ;
- k_{ij} = keuntungan transaksi ke- j pada kategori ke- i ; dan
- n_i = jumlah transaksi dalam kategori ke- i .

f. Visualisasi data

Hasil analisis disajikan menggunakan diagram batang, diagram lingkaran, histogram, dan visualisasi lain yang sesuai. Diagram batang digunakan untuk membandingkan tingkat pembelian, keuntungan, dan kontribusi setiap cabang. Visualisasi berdasarkan gender digunakan untuk menunjukkan perbandingan jumlah pembelian konsumen laki-laki dan perempuan. Apabila digunakan, *dashboard* Tableau menyajikan hasil analisis secara terpadu dan interaktif.

g. Interpretasi dan penyusunan rekomendasi

Tahap terakhir dilakukan dengan menginterpretasikan pola yang diperoleh dari proses EDA. Kategori dengan frekuensi pembelian tertinggi dianggap sebagai produk yang paling diminati, sedangkan kategori dengan akumulasi keuntungan terbesar ditetapkan sebagai kategori yang paling menguntungkan. Cabang atau kota dengan total transaksi terbesar diidentifikasi sebagai wilayah dengan kontribusi penjualan tertinggi. Perbandingan berdasarkan gender digunakan untuk mendukung segmentasi pelanggan.

Temuan tersebut kemudian diterjemahkan menjadi rekomendasi bisnis mengenai pengelolaan persediaan, prioritas kategori produk, strategi promosi, evaluasi kinerja cabang, dan penentuan target pemasaran. Dengan demikian, hasil EDA tidak hanya menjelaskan kondisi data penjualan, tetapi juga dapat menjadi dasar pendukung pengambilan keputusan untuk meningkatkan pendapatan Supermarket XYZ.

3. ANALISI DAN PEMBAHASAN

3.1 Implementasi *Exploratory Data Analysis*

Penerapan *Exploratory Data Analysis* (EDA) dilakukan untuk memperoleh gambaran menyeluruh mengenai karakteristik data transaksi Supermarket XYZ. Proses analisis menggunakan bahasa pemrograman Python dengan bantuan pustaka NumPy dan Pandas untuk pengolahan data, sedangkan Matplotlib dan Seaborn digunakan untuk visualisasi. Selain itu, Tableau dimanfaatkan untuk menyajikan hasil analisis dalam bentuk *dashboard* yang lebih interaktif dan mudah dipahami.

Analisis diarahkan pada empat tujuan utama, yaitu mengidentifikasi kategori produk yang paling banyak dibeli, menentukan kategori dengan keuntungan terbesar, mengetahui kota atau cabang dengan jumlah penjualan tertinggi, dan menganalisis pembelian berdasarkan gender konsumen. Analisis tambahan berupa statistik deskriptif, nilai *skewness*, pemeriksaan *outlier*, dan pengujian perbedaan jumlah pembelian berdasarkan gender juga dilakukan untuk memperkuat interpretasi hasil.

3.2 Prapemrosesan Data

a. Pemuatan dan Pemeriksaan Data

Dataset transaksi penjualan dimuat menggunakan fungsi pengolahan data pada pustaka Pandas. Berdasarkan hasil pemeriksaan awal, dataset terdiri atas 1.000 baris dan 17 kolom. Setiap baris merepresentasikan satu transaksi, sedangkan kolom memuat atribut seperti *Invoice ID*, *Branch*, *City*, *Customer Type*, *Gender*, *Product Line*, *Unit Price*, *Quantity*, *Tax*, *Total*, *Date*, *Payment*, *cogs*, *Gross Margin Percentage*, *Gross Income*, dan *Rating*.

Pemeriksaan menggunakan fungsi `info()` menunjukkan bahwa atribut kategoris telah disimpan dalam tipe data objek, sedangkan atribut perhitungan penjualan menggunakan tipe data numerik berupa *integer* dan *float*. Kesesuaian tipe data tersebut memudahkan proses pengelompokan, perhitungan statistik, dan visualisasi.

b. Pemeriksaan *Missing Value*

Pemeriksaan nilai kosong dilakukan menggunakan fungsi `isnull().sum()`. Hasilnya menunjukkan bahwa seluruh atribut memiliki nilai *missing value* sebesar nol. Dengan demikian, sebanyak 1.000 data transaksi dapat digunakan dalam analisis tanpa melakukan penghapusan baris ataupun imputasi.

Ketiadaan nilai kosong menunjukkan bahwa dataset memiliki kelengkapan yang baik. Meskipun demikian, data tetap diperiksa dari kemungkinan duplikasi, ketidaksesuaian tipe data, dan nilai ekstrem agar hasil analisis tidak mengalami distorsi. Tahap ini penting karena kualitas hasil EDA sangat bergantung pada konsistensi data yang digunakan.

3.3 Analisis Kategori Produk yang Paling Banyak Dibeli

Hasil pengelompokan jumlah produk berdasarkan *product line* menunjukkan bahwa *Electronic Accessories* merupakan kategori yang paling banyak dibeli dengan kontribusi sebesar 17,62%. Posisi berikutnya ditempati oleh *Food and Beverages* sebesar 17,28%, *Sports and Travel* sebesar 16,70%, *Home and Lifestyle* sebesar 16,53%, *Fashion Accessories* sebesar 16,37%, serta *Health and Beauty* sebesar 15,50%.

Tabel 1. Analisis Kategori Produk yang Paling Banyak Dibeli

Kategori Produk	Persentase Pembelian
<i>Electronic Accessories</i>	17,62%
<i>Food and Beverages</i>	17,28%
<i>Sports and Travel</i>	16,70%
<i>Home and Lifestyle</i>	16,53%
<i>Fashion Accessories</i>	16,37%
<i>Health and Beauty</i>	15,50%

Dominasi *Electronic Accessories* menunjukkan tingginya permintaan konsumen terhadap produk elektronik dan aksesoris pendukung. Namun, selisih antarkategori relatif kecil, yaitu hanya 2,12 poin persentase antara kategori tertinggi dan terendah. Kondisi tersebut memperlihatkan bahwa pembelian konsumen tersebar cukup merata dan Supermarket XYZ tidak hanya bergantung pada satu kategori produk.

Berdasarkan temuan tersebut, perusahaan dapat menjaga ketersediaan *Electronic Accessories* sebagai kategori dengan jumlah pembelian tertinggi. Pada saat yang sama, pengelolaan kategori lain tetap perlu dipertahankan karena masing-masing memberikan kontribusi yang relatif seimbang. Strategi promosi dapat dilakukan melalui pemberian diskon, penempatan produk pada area yang mudah terlihat, dan penawaran paket produk yang saling melengkapi.

3.4 Analisis Kategori Produk dengan Keuntungan Tertinggi

Hasil agregasi keuntungan menunjukkan bahwa kategori *Food and Beverages* memberikan keuntungan terbesar dengan nilai sekitar 51.000 berdasarkan satuan nilai dalam dataset. Kategori tersebut diikuti oleh *Sports and Travel* dengan keuntungan sekitar 50.000. Selanjutnya, *Electronic Accessories*, *Fashion Accessories*, dan *Home and Lifestyle* menghasilkan keuntungan pada kisaran 48.000–49.000, sedangkan *Health and Beauty* mencatat keuntungan terendah sekitar 44.000–45.000.

Tingginya keuntungan kategori *Food and Beverages* dapat dikaitkan dengan sifat produknya sebagai barang konsumsi yang dibutuhkan secara rutin. Permintaan yang berulang menciptakan peluang bagi perusahaan untuk mempertahankan perputaran produk dan memperoleh keuntungan secara konsisten. Supermarket XYZ dapat mengoptimalkan kategori ini melalui pengendalian persediaan, penambahan variasi produk, promosi berkala, dan strategi *bundling*.

Kategori *Health and Beauty* memerlukan evaluasi karena menghasilkan keuntungan paling rendah. Evaluasi dapat mencakup kesesuaian harga, variasi produk, penempatan barang, efektivitas promosi, dan tingkat perputaran persediaan. Akan tetapi, kategori tersebut tidak seharusnya langsung dikurangi hanya berdasarkan nilai keuntungan agregat. Manajemen perlu

mempertimbangkan margin per produk, biaya persediaan, dan kontribusinya terhadap kebutuhan konsumen sebelum mengambil keputusan.

Hasil ini juga menunjukkan bahwa kategori yang paling banyak dibeli belum tentu menghasilkan keuntungan tertinggi. *Electronic Accessories* memiliki jumlah pembelian terbesar, tetapi *Food and Beverages* menghasilkan keuntungan paling tinggi. Dengan demikian, keputusan mengenai persediaan dan promosi perlu mempertimbangkan jumlah penjualan sekaligus keuntungan yang diperoleh.

3.5 Analisis Pembelian Berdasarkan Gender

Hasil analisis menunjukkan bahwa konsumen perempuan menyumbang 50,10% dari jumlah pembelian, sedangkan konsumen laki-laki menyumbang 49,90%. Selisih sebesar 0,20 poin persentase menunjukkan bahwa komposisi pembelian berdasarkan gender hampir seimbang.

Tabel 2. Analisis Pembelian Berdasarkan Gender

Gender	Persentase pembelian
Perempuan	50,10%
Laki-laki	49,90%

Komposisi yang hampir seimbang menunjukkan bahwa produk Supermarket XYZ memiliki jangkauan pasar yang luas bagi kedua kelompok konsumen. Oleh karena itu, perusahaan sebaiknya tidak hanya memusatkan kegiatan pemasaran pada satu gender. Strategi pemasaran dapat disusun berdasarkan preferensi kategori produk pada masing-masing kelompok agar promosi menjadi lebih relevan.

Pengujian perbedaan rata-rata kuantitas pembelian konsumen perempuan dan laki-laki dilakukan menggunakan *independent two-sample t-test*. Hasil pengujian memperoleh nilai *p-value* sebesar 0,010851, lebih kecil daripada taraf signifikansi 0,05. Oleh karena itu, hipotesis nol ditolak sehingga terdapat perbedaan rata-rata kuantitas pembelian yang signifikan secara statistik antara konsumen perempuan dan laki-laki.

Hasil tersebut perlu dibedakan dari persentase agregat pembelian yang hampir seimbang. Persentase 50,10% dan 49,90% menunjukkan pembagian total kuantitas pembelian, sedangkan uji *t* membandingkan rata-rata kuantitas pada tingkat transaksi. Dengan demikian, total kontribusi kedua gender dapat terlihat hampir sama, tetapi pola rata-rata pembeliannya tetap berbeda secara statistik. Temuan ini dapat menjadi dasar untuk melakukan analisis lanjutan mengenai kategori produk yang paling diminati oleh setiap gender.

3.6 Analisis Distribusi Data

a. Analisis *Skewness*

Analisis *skewness* dilakukan terhadap variabel *Quantity*, *Gross Income*, dan *Total*. Hasil perhitungan ditampilkan pada tabel berikut.

Tabel 3. Analisis *Skewness*

Variabel	Nilai <i>skewness</i>	Interpretasi
<i>Quantity</i>	0,012491	Hampir simetris
<i>Gross Income</i>	0,892570	Miring ke kanan
<i>Total</i>	0,892570	Miring ke kanan

Nilai *skewness* variabel *Quantity* yang mendekati nol menunjukkan bahwa kuantitas pembelian memiliki distribusi yang hampir simetris. Sementara itu, *Gross Income* dan *Total*

memiliki kemencengan positif. Hal tersebut menunjukkan bahwa sebagian besar transaksi berada pada nilai rendah hingga menengah, sedangkan sejumlah transaksi bernilai tinggi membentuk ekor distribusi di sebelah kanan.

Kesamaan nilai *skewness* pada *Gross Income* dan *Total* menunjukkan pola distribusi yang serupa. Hal ini wajar karena keuntungan kotor pada dataset memiliki hubungan perhitungan dengan nilai transaksi. Distribusi yang miring ke kanan juga menjelaskan mengapa rata-rata *gross income* lebih tinggi dibandingkan mediannya.

b. Analisis Outlier

Pemeriksaan *outlier* pada variabel *Gross Income* dilakukan menggunakan metode *Interquartile Range* (IQR). Hasil perhitungan memperoleh batas bawah sebesar -18,86 dan batas atas sebesar 47,23. Sebanyak 0,90% data berada di luar batas tersebut.

Proporsi *outlier* yang rendah menunjukkan bahwa sebagian besar transaksi berada dalam rentang yang wajar. Meskipun demikian, data ekstrem tidak sebaiknya langsung dihapus karena dapat merepresentasikan transaksi dengan jumlah pembelian atau nilai transaksi yang benar-benar tinggi. Oleh karena itu, nilai tersebut perlu diverifikasi terlebih dahulu. Jika terbukti valid, transaksi ekstrem dapat dipertahankan dan dianalisis sebagai segmen transaksi bernilai tinggi yang berpotensi memberikan informasi bisnis.

3.7 Analisis Statistik Deskriptif

Analisis statistik deskriptif terhadap *Gross Income* dilakukan pada 1.000 transaksi. Hasilnya menunjukkan nilai rata-rata sebesar 15,379369 dan median sebesar 12,088000. Nilai minimum tercatat sebesar 0,508500, sedangkan nilai maksimum mencapai 49,650000. Simpangan baku sebesar 11,708825 menunjukkan adanya variasi keuntungan antartransaksi.

Tabel 4. Analisis Statistik Deskriptif

Statistik	Nilai <i>Gross Income</i>
Jumlah data	1.000
Rata-rata	15,379369
Simpangan baku	11,708825
Minimum	0,508500
Kuartil pertama	5,924875
Median	12,088000
Kuartil ketiga	22,445250
Maksimum	49,650000

Rata-rata yang lebih tinggi daripada median memperkuat hasil analisis *skewness* bahwa distribusi *Gross Income* cenderung miring ke kanan. Sebanyak 50% transaksi berada pada kisaran 5,924875 hingga 22,445250. Sementara itu, sebagian kecil transaksi dengan keuntungan tinggi menaikkan nilai rata-rata keseluruhan.

Temuan ini dapat digunakan untuk menyusun strategi peningkatan nilai transaksi. Pada transaksi dengan keuntungan rendah, perusahaan dapat menerapkan *cross-selling*, paket produk, atau batas minimum pembelian untuk memperoleh promosi. Untuk transaksi bernilai tinggi, perusahaan dapat mengembangkan program loyalitas dan penawaran khusus agar pelanggan terdorong melakukan pembelian ulang.

3.8 Pembahasan Penelitian

Penerapan EDA berhasil mengidentifikasi empat informasi utama yang berpotensi mendukung peningkatan pendapatan Supermarket XYZ. Pertama, *Electronic Accessories* menjadi kategori dengan jumlah pembelian tertinggi sehingga ketersediaan dan promosinya perlu dipertahankan. Kedua, *Food and Beverages* memberikan keuntungan terbesar sehingga dapat diprioritaskan dalam strategi pengelolaan produk. Ketiga, pembelian konsumen perempuan dan laki-laki secara agregat hampir seimbang, tetapi rata-rata kuantitas pembeliannya menunjukkan perbedaan yang signifikan.

Hasil tersebut menunjukkan bahwa strategi peningkatan pendapatan tidak cukup hanya didasarkan pada jumlah produk yang terjual. Perusahaan juga perlu mempertimbangkan keuntungan, kinerja wilayah, dan karakteristik konsumen. Penggabungan indikator tersebut memungkinkan manajemen menyusun strategi persediaan, promosi, dan segmentasi pelanggan secara lebih terarah.

Meskipun demikian, EDA pada dasarnya bersifat eksploratif sehingga hasilnya menggambarkan pola dan keterkaitan dalam data, bukan membuktikan hubungan sebab-akibat. Oleh karena itu, istilah “faktor peningkatan pendapatan” dalam penelitian ini merujuk pada indikator yang berpotensi mendukung peningkatan pendapatan berdasarkan pola data. Analisis lanjutan dengan variabel biaya, periode promosi, harga, diskon, dan kondisi operasional diperlukan untuk mengukur pengaruh setiap faktor secara lebih mendalam.

4. KESIMPULAN

Penerapan metode *Exploratory Data Analysis* (EDA) pada data penjualan Supermarket XYZ berhasil mengidentifikasi pola produk, keuntungan, wilayah pemasaran, dan karakteristik konsumen. Kategori *Electronic Accessories* menjadi produk yang paling diminati dengan persentase pembelian sebesar 17,62%. Kategori *Food and Beverages* menghasilkan keuntungan tertinggi sekitar Rp50.000, sedangkan *Health and Beauty* memperoleh keuntungan terendah sekitar Rp10.000. Berdasarkan gender, konsumen perempuan melakukan pembelian lebih banyak daripada konsumen laki-laki, meskipun perbedaannya relatif kecil. Hasil analisis ini dapat menjadi dasar bagi Supermarket XYZ dalam menyusun strategi pemasaran, pengelolaan persediaan, dan segmentasi pelanggan untuk mendukung peningkatan pendapatan.

REFERENCES

- Abdurrahman, G. (2017). Analisis Aturan Asosiasi Data Transaksi Supermarket Menggunakan Algoritma Apriori. *Jurnal Sistem & Teknologi Informasi Indonesia*, 2(2), 100–111.
- Aprizal, Hasriani, dan Ningsih, W. (2016). Implementasi Data Mining untuk Penentuan Posisi Barang pada Rak Menggunakan Metode Apriori pada PT Midi Utama Indonesia. *Techno.com*, 15(4), 335–342.
- Listriani, D., Setyaningrum, A. H., dan Eka, F. (2016). Penerapan Metode Asosiasi Menggunakan Algoritma Apriori pada Aplikasi Analisis Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro). *Jurnal Teknik Informatika*, 9(2), 120–127.
- Maharani et al. (2017). Implementasi Data Mining untuk Pengaturan Layout Minimarket dengan Menerapkan Association Rule. *Jurnal Riset Komputer*, 4(4), 6–11. <https://doi.org/10.30865/jurikom.v4i4.686>
- Nelisa dan Hadi, A. F. (2018). Perancangan Aplikasi Data Mining Transaksi Penjualan untuk Mengetahui Pola Beli Konsumen pada Toko Singgalang Padang Menggunakan Algoritma Apriori Berbasis Web. *Majalah Ilmiah UPI YPTK*, 25(1), 37–44.
- Nurdin dan Astika, D. (2015). Penerapan Data Mining untuk Menganalisis Penjualan Barang pada Supermarket Sejahtera Lhokseumawe. *Techsi: Jurnal Penelitian Teknik Informatika*, 6(1), 134–155.
- Rahmawati, Nasution, Y. N., dan Amijaya, F. D. T. (2017). Aplikasi Data Mining Market Basket Analysis untuk Menemukan Pola Pembelian di Toko Metro Utama Balikpapan. *Jurnal Eksponensial*, 8(1), 1–8.
- Santoso, H., Hariyadi, I. P., dan Prayitno. (2016). Data Mining Analisis Pola Pembelian Produk. *Seminar Nasional Teknologi Informasi dan Multimedia*, 19–24.
- Sandi, A. P. (2024). Implementasi Data Mining sebagai Penentu Persediaan Produk dengan Algoritma FP-Growth pada Data Penjualan Sinarmart.
- Susanto, S., dan Suryadi, D. (2010). *Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: Andi.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley Publishing Company.



Wulandari, R. T. (2017). Data Mining: Teori dan Aplikasi RapidMiner. Yogyakarta: Gava Media.
Zai, C. (2022). Implementasi Data Mining sebagai Pengolahan Data. Portaldata, 2(3), 1–12.